The practice of data base administration

by JAY-LOUISE WELDON
New York University
New York, New York

The position or role of data base administrator has been described and discussed since the earliest specifications for data base management systems. Most of the literature on DBA is normative—describing in detail what the DBA function should include. Different authors have focused on the DBA's role in introducing the data base concept to organizations, the functions that the DBA should perform, and how to establish and perform the DBA's responsibilities within an organization. In practice, however, the organization and content of the DBA function may be quite different from the ideal. Other factors related to the DBA's technical and organizational environment may influence the DBA's role. This paper summarizes the results of an exploratory survey on data base administration which was undertaken as a first step toward identifying and understanding these factors.

METHODOLOGY

The survey included 25 DBA groups within firms located in the New York metropolitan area. Individual interviews were arranged with the manager of each DBA group. The interviews were unstructured and the questions generally open-ended. However, the interviewer did follow an interview outline that included questions on three topic areas: 1) basic characteristics of the firm, 2) organizational characteristics of the firm and of the DBA group, and 3) the tasks performed by the DBA group. The interviews were summarized immediately after they occurred and the responses were later coded according to previously-defined classification schemes.

Based on existing literature on DBA, seven characteristics were defined as independent variables, i.e. factors that were thought to influence DBA organization and content:

| Age of the DBA organization | The number of years that the DBA organization/position had existed. |
| Origin relative to DBMS | Was DBA instituted before, with, or after the DBMS? |
| DBMS | Which DBMS package was used? |

| Industry | The industry most descriptive of the firm's business operations. |
| Installation size | An index combining the size of the hardware and the number of bytes stored in data bases. |
| EDP organization | The degree of centralization of the EDP organization. |
| Data base scope | An index based on the number of different application areas with current or planned data base support. |

Five factors descriptive of the DBA function were also defined for use as dependent variables:

| Size | Size of DBA staff with respect to total EDP staff. |
| Organizational position of DBA | Described by the number of levels between the DBA and the head of EDP, the position of DBA with respect to its primary users, and an indicator of organizational change since DBA's inception (higher, lower, no change). |
| Orientation of DBA staff | Ratings of the DBA staff on a technical-administrative scale and an application-systems scale. |
| DBA organizational structure | DBA's span of control and the number of organizational levels below DBA. |
| DBA tasks | A vector of yes-no indicators for 48 representative DBA tasks. |

Responses were tabulated for each of the variables above to obtain descriptive statistics (frequency distributions, means or medians, etc.) on each. The task indicators were summed for the sample and used to rank order the DBA tasks with respect to frequency of mention.
To explore possible relationships among the independent and dependent variables, comparative statistics (Chi-square tests, t-tests, and rank-order correlations) were used. The sample was partitioned by each of the seven independent variables and values of the dependent variables for the subsamples were compared.

RESULTS*

The nature of the survey respondents can be determined from the descriptive statistics on the independent variables. The sample was largely self-selected, and therefore not demonstrably random. However, the characteristics show a reasonably diverse group, representing a spectrum of age (one to five years), industry, installation size, and data base scope. One DBMS (IMS) dominates the sample (used by 65 percent of the respondents) as does one mode of EDP organization (75 percent were centralized).

The characteristics of the DBA groups can be summarized as follows:

- **Size**—The size of the DBA group relative to EDP staff size ranged from 0.4 percent to 12.5 percent with a mean of 3.34 percent and a standard deviation of 3.01 (actual staff sizes ranged from one to 28). If projected staff increases are included, the mean increases to 3.86 percent.

- **Organizational position**—Organizational position was of interest since most authorities claim that DBA requires a high management level for success. Most of the groups surveyed were two or more levels below the top EDP manager. The range for this variable was one to four, with only 25 percent reporting directly to the top EDP manager. Most of the DBA groups had changed their organizational position since their inception, most for the better. This suggests that DBA might start low in the organization and gain in position as the function matures. Over half of the groups, however, were placed on a lower organizational level than their primary users (e.g. applications programmers). This suggests that DBA is still considered a support function by most EDP groups. Only 10 percent of the DBA managers were on a higher level than the manager of their primary user group.

- **Staff Orientation**—Each DBA group was assigned a rating on two five-point scales according to the job titles or descriptions provided for their staff members. (Scale 1: 1 = All administrative, no technical; 2 = Mixed; 3 = All technical. Scale 2: 1 = Applications-oriented; 2 = Mixed; 4 = Systems-oriented.) The tabulation of the first scale showed a clear dichotomy, with twice as many groups rated technical as rated administrative (mean rating = 3.55, standard deviation = 1.5). Responses on Scale 2 showed that most groups were applications-oriented (45 percent) or mixed (40 percent), and none were entirely systems-oriented. (Mean ratings = 2.4, standard deviation = 1.1).

- **Organizational Structure**—Almost all of the respondents (89 percent) had either one level (42 percent) or two levels (47 percent) of staff reporting to the DBA manager. For both flat and pyramid structures, the most common span of control was three (the range was zero to six). The most common functional designations associated with these three subgroups were: Support for data base design and maintenance, data standards (including data dictionary), and DBMS support.

- **Tasks**—The tasks performed by each DBA group were recorded on a menu of 48 representative tasks culled from the literature.** By summing across the sample for each task, the proportion of DBA groups performing each task was determined and the tasks were ranked accordingly. Table I shows the tasks most, and least, frequently mentioned.

FACTORS INFLUENCING DBA CHARACTERISTICS

Relationships were detected between each of the independent variables and one or more DBA characteristics. While some of the other cross-classifications suggested possible relationships, due to the small sample size (N = 20, five responses being unacceptable for various reasons), they were not statistically significant. Except where noted, the following discussion will be limited to significant relationships (p<.05).

The length of time that a DBA had been in existence was found to be related to four of the five DBA characteristics. Only relative staff size was not significantly related. Inspection of the data shows, however, that the younger groups (three years or less) show more variation in size and a higher proportion of small staff sizes than the older groups.

As might be expected, the older groups show more internal structure and have experienced more organizational change than the younger groups. Younger groups also tend to be more applications-oriented, while older groups are more systems-oriented. This might be associated with the development cycle of the data base applications (i.e. younger—design, older—operational).

DBA appears to start as a technical support group for one or more applications areas (and below them in the organization). Over time DBA moves to a higher level (equal to the primary user group) functioning then in a more consultative fashion.

One interesting, though non-significant, result suggests that a change may have occurred in this pattern over the past two years. The youngest DBA groups (two years or less) reported a broad range of applications supported, similar to the oldest groups (greater than four years). This, coupled with the fact that none of the young groups supported only one application area, suggests that DBA may now be avoiding the project approach, i.e. supporting only one application area, which was common in the past.

* An extended discussion of the survey results and supporting data can be found in another paper by the author.*

** The full list of tasks is shown in the Appendix.
TABLE I—DBA Tasks Mentioned by Survey Respondents

<table>
<thead>
<tr>
<th>Most Frequently Performed</th>
<th>%</th>
<th>Least Frequently Performed</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td></td>
<td>Task</td>
<td></td>
</tr>
<tr>
<td>Maintain Data Dictionary System</td>
<td>75</td>
<td>Select DB-related hardware</td>
<td>0</td>
</tr>
<tr>
<td>Evaluate DB software</td>
<td>60</td>
<td>Enforce DB retention policies</td>
<td>0</td>
</tr>
<tr>
<td>Maintain DB descriptions</td>
<td>60</td>
<td>Requirement analysis for DB applications</td>
<td>5</td>
</tr>
<tr>
<td>Reorganize DBs</td>
<td>60</td>
<td>Schedule computer time</td>
<td>5</td>
</tr>
<tr>
<td>Recover DBs</td>
<td>60</td>
<td>Determine program structure</td>
<td>10</td>
</tr>
<tr>
<td>Select DB software</td>
<td>55</td>
<td>Evaluate DB-related hardware</td>
<td>15</td>
</tr>
<tr>
<td>Forecast DB growth</td>
<td>55</td>
<td>Design forms/procedures for DB applications</td>
<td>15</td>
</tr>
<tr>
<td>Generate DB descriptions</td>
<td>55</td>
<td>Enforce standards for application coding</td>
<td>15</td>
</tr>
<tr>
<td>Monitor DB performance</td>
<td>55</td>
<td>Maintain DC monitor</td>
<td>15</td>
</tr>
<tr>
<td>Develop standards for data element names</td>
<td>55</td>
<td>Enforce standards for documentation</td>
<td>20</td>
</tr>
<tr>
<td>Determine physical structures</td>
<td>50</td>
<td>Load DBs</td>
<td>50</td>
</tr>
<tr>
<td>Specify DB access policies</td>
<td>50</td>
<td>Set policies on DB backup and recovery</td>
<td>50</td>
</tr>
<tr>
<td>Load DBs</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set policies on DB backup and recovery</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DBA groups established before a DBMS was installed tend to be larger (in relative size) than those instituted with or after the DBMS. This, however, may be indirectly related to age, since groups started before the DBMS are older than most of the groups started with or after the DBMS.

Installation size was found to be related to the level of the DBA with respect to the top EDP manager and also to the amount of organization change experienced by the DBA. DBAs in organizations with large installations (large model hardware and large data bases) tend to be lower in the EDP organization than those in smaller installations. These DBAs also reported more organizational change (both positive and negative) in the DBA group than did DBAs at smaller installations.

The relationship between the DBMS package used and DBA characteristics was significant for only two variables: DBA's span of control and DBA tasks. Both of these results compared IMS DBAs with DBAs using other DBMS packages. IMS DBAs were found to have larger span of control, i.e. more persons reporting to them, than the other DBAs. Further, the ranks assigned to DBA tasks by the IMS DBAs were found to be inversely correlated with those assigned by the other DBAs. The IMS DBAs emphasize tasks related to planning and control (forecast growth, data dictionary, data name standards) while the others emphasize operational support tasks (e.g. data base load, troubleshooting).

In contrast with the sparse relationships detected between DBA organizational characteristics and the independent variables, the content of the DBA function was related to all but one. Differences in the rank order of tasks performed were detected among groups partitioned by every independent variable, except installation size. For a fuller discussion, see Reference 9.

CONCLUSIONS

The results of this survey support two major conclusions. First, the organizational aspects of data base administration groups are affected primarily by the length of time that the group has existed. This suggests a maturation process for DBA, perhaps similar to Nolan's stages of EDP growth. To formulate such a hypothesis, criteria defining each stage in the process must be specified and characteristics of DBA organization and structure related to each stage.

Second, it appears that the tasks performed by DBA groups vary independently from organizational structure or position. The composition of a DBA's job and the rank ordering of tasks within it are influenced by several factors, including the DBMS package used, EDP organization type, and the scope of data base applications. Additional task data from stratified random samples for each of the related variables could be used to explore these relationships in greater detail. It may then be possible to develop task profiles for the several different types of data base administration functions that exist in practice.

APPENDIX—CLASSIFICATION OF DBA TASKS

Planning and Management
- Evaluate data base software (e.g. DBMS, DDDS)
- Select data base software
- Evaluate data base-related hardware (e.g. disks, terminals)
- Select data base-related hardware
- Define implementation strategy for data bases
- Forecast data base growth
- Set operational goals (performance, downtime)
- Set application priorities
- Hire, fire, promote data base personnel

Data Base Design
- Requirement analysis for data base applications (data identification)
- Develop data definitions
- Determine data structures (views)
- Determine physical structures
- Generate data base descriptions (DDL)
- Design integrity controls for data base applications
- Design forms and procedures for data base applications
Application Programming/Testing

Determine program structures (processes)
Develop standards for application coding
Enforce standards for application coding
Application system testing

Controls

Generate data base descriptions (DDL)
Maintain data base descriptions
Design integrity controls for data base applications
Specify data base access policies
Record data base usage
Monitor data base controls
Develop standards for data names
Develop standards for application coding
Develop standards for documentation
Maintain DDS
Set data base retention policies
Select/design security techniques (passwords, locks, etc.)

Operational Support

Monitor data base controls
Load data bases
Reorganize data bases
Recover data bases
Monitor data base performance
Tune data base to meet operational goals
Troubleshooting (i.e. track down problems)
Enforce standards for data element naming
Enforce standards for application coding
Enforce standards for documentation
Set policies on data base backup and recovery
Schedule computer time (e.g. when data base is up)
Develop data base utilities (data compression, encryption, query languages)
Enforce data base retention policies

DBMS Support

Install new DBMS features
Modify or fix DBMS
Maintain data communications monitor
Develop data base utilities (data compression, encryption, query languages)

User Support

Prepare data base documentation
Disseminate data base documentation
Educate data base users
Maintain DDDS
Maintain query language(s)

REFERENCES